

“Express Mail” Mailing Label No. **EL960827955US**

PATENT APPLICATION
ATTORNEY DOCKET NO. SUN-P9610-SPL

INTEGRATED CIRCUIT ASSEMBLY MODULE
THAT SUPPORTS CAPACITIVE
COMMUNICATION BETWEEN
SEMICONDUCTOR DIES

Inventors: Ivan E. Sutherland, Robert J. Drost,
Gary R. Lauterbach, and Howard L. Davidson

RELATED APPLICATION

[0001] This application hereby claims priority under 35 U.S.C. 119 to U.S. Provisional Patent Application No. 60/443,808, filed on 29 January 2003, entitled “Silicon Module for Face to Face Chips,” by inventor Robert J. Drost (Attorney Docket No. SUN-P9610PSP).

GOVERNMENT LICENSE RIGHTS

[0002] This invention was made with United States Government support under Contract No. NBCH020055 awarded by the Defense Advanced Projects Administration. The United States Government has certain rights in the invention.

BACKGROUND

Field of the Invention

[0003] The present invention relates to semiconductor integrated circuits. More specifically, the present invention relates to an integrated circuit assembly module that supports capacitive communication between semiconductor dies.

Related Art

[0004] Integrated circuit chips ordinarily communicate with one another through external wiring on a printed circuit board. In order to adapt the tiny dimensions of the integrated circuit to the larger dimensions of the wires on the printed circuit board, the integrated circuit is typically mounted in a “package” made of plastic or ceramic.

[0005] Integral to the package are a set of metal conductors, one for each connection to the integrated circuit. These metal conductors are typically connected to the integrated circuit via gold or aluminum “bonding wires.” These bonding wires are typically 25 microns in diameter, and are attached to “bonding pads” on the integrated circuit that typically measure 100 microns square. The other end of the bonding wires are attached to metal conductors that range from 200 to 500 microns in size, with the larger ends attached to a printed circuit board. In order for signals to travel from one integrated circuit to another, they must pass through several sets of drivers to drive the signal over increasingly larger areas.

[0006] As the trend towards smaller and smaller integrated circuits with increasing numbers of transistors continues, limited communications paths on to and off of the chip are quickly becoming a major bottleneck to processor performance. Because advances in packaging technologies have not kept pace with increasing semiconductor integration densities, inter-chip communication

bandwidth has not kept pace with on-chip computational power. Moreover, as feature sizes continue to decrease, the delay for communications off of the chip has increased, because more time is required to amplify small on-chip signal lines to drive large bonding pads.

[0007] Hence, what is needed is a method and an apparatus for communicating between semiconductor chips in a manner that reduces the amount of space and power required for the communication process.

SUMMARY

[0008] One embodiment of the present invention provides an integrated circuit assembly module, including a first semiconductor die with an active face upon which active circuitry and signal pads reside and a back face opposite the active face, and a second semiconductor die with an active face upon which active circuitry and signal pads reside and a back face opposite the active face. The first semiconductor die and the second semiconductor die are positioned face-to-face within the assembly module so that signal pads on the first semiconductor die overlap with signal pads on the second semiconductor die, thereby facilitating capacitive communication between the first semiconductor die and the second semiconductor die. Additionally, the first semiconductor die and the second semiconductor die are pressed together between a first substrate and a second substrate so that a front side of the first substrate is in contact with the back face of the first semiconductor die and a front side of the second substrate is in contact with the back face of the second semiconductor die.

[0009] In a variation on this embodiment, the integrated circuit assembly module further includes a first heat removal mechanism coupled to the back side of the first substrate opposite the front side, and a second heat removal mechanism coupled to the back side of the second substrate opposite the front side.

[0010] In a further variation, the first and second heat removal mechanisms include cooling fins to facilitate the transfer of heat to air passing across the cooling fins.

[0011] In a further variation, the first and second heat removal mechanisms include graphite foam or metal with fin-like structures which facilitate in the transfer of heat to a liquid pumped through the graphite foam or metal.

[0012] In a variation on this embodiment the integrated circuit assembly module further includes a first power supply coupled to the back side of the first substrate, and a second power supply coupled to the back side of the second substrate.

[0013] In a further variation, the first and second substrates include metal layers that facilitate supplying power to the first and second semiconductor dies.

[0014] In a further variation, the first and second substrates include power connectors on the back sides of the first and second substrates.

[0015] In a further variation, the integrated circuit assembly module further comprises Micro Electro-Mechanical System (MEMS) spring contacts that provide power from the metal layers within the first and second substrates to the first and second semiconductor dies. In this embodiment, a first set of MEMS spring contacts on the front side of the first substrate contact the active side of the second semiconductor die, and a second set of MEMS spring contacts on the front side of the second substrate contact the active side of the first semiconductor die.

[0016] In a further variation, a first power regulator is incorporated within the first substrate, and a second power regulator is incorporated within the second substrate.

[0017] In a further variation, the integrated circuit assembly module further includes an I/O semiconductor die with an active face upon which active

circuitry and signal pads reside, and a back face opposite the active face. This I/O semiconductor die and the second semiconductor die are arranged face-to-face such that signal pads on the I/O semiconductor die overlap with signal pads on the second semiconductor die, thereby facilitating capacitive communication between the I/O semiconductor die and the second semiconductor die. Additionally, the I/O semiconductor die is located on an edge of the first substrate to facilitate communications into and out of the integrated circuit assembly module. The edge of the first substrate extends beyond the edge of the second substrate so that a portion of the active face of the I/O semiconductor die is exposed to facilitate external connections.

[0018] In a further variation, the integrated circuit assembly module further includes optical connection pads located on the exposed portion of the active side of the I/O semiconductor die.

[0019] In a further variation, the integrated circuit assembly module further includes electrical connection pads located on the exposed portion of the active side of the I/O semiconductor die.

BRIEF DESCRIPTION OF THE FIGURES

[0020] FIG. 1 illustrates an integrated circuit assembly module in accordance with an embodiment of the present invention.

[0021] FIG. 2 illustrates a portion of an integrated circuit assembly module with power connections in accordance with an embodiment of the present invention.

[0022] FIG. 3 presents proximity communication overlap patterns in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION

[0023] The following description is presented to enable any person skilled in the art to make and use the invention, and is provided in the context of a particular application and its requirements. Various modifications to the disclosed embodiments will be readily apparent to those skilled in the art, and the general principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the present invention. Thus, the present invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features disclosed herein.

Background

[0024] In one embodiment of the present invention, an integrated circuit assembly module includes an almost symmetric stacking of components. An arbitrary number of chips are tiled between two substrates (wafers). This tiling arrangement does not need to be uniform. Hence, the integrated circuit assembly module could contain chips of different sizes and could contain empty positions.

[0025] The wafers contain alignment structures that precisely position chips that are placed between them. Two layers of chips are used in this system, one layer for each of the two wafers. Each layer of chips sits with the back-side of the chips against the wafer, and the front side (the side that contains the active circuitry and wires) facing the chips on the other wafer.

[0026] Each wafer may contain alignment structures for one layer of chips. Alternately, one wafer may contain the alignment structures for both layers of chips. If both wafers contain alignment structures, which may be simpler to fabricate, then it is required that the wafers be aligned to each other in order to align the relative positions of the chips between the two layers. If one wafer

contains the alignment structures for both chips, then chips may be more accurately aligned.

[0027] The alignment structures may include posts built up on the flat wafer surfaces. The chips may be aligned by providing a cavity using a partial or full perimeter of posts. Alternately, the chips may be aligned by posts that fit through holes in the chips. Two holes per chip would be required, and more may be used to add redundancy. Aligning using a cavity requires less chip complexity, but it also requires that the chips be trimmed to a precise size to fit the cavities. Aligning using posts and holes may provide greater accuracy because the holes on the chip could be created by etching the holes and using an annulus formed by the existent metal interconnect layers to guide the etching process.

[0028] The chips do not need to be permanently attached to the wafers. This makes it possible to remove a chip from a fully assembled integrated circuit assembly module if it is determined that the chip is defective. Additionally, chips may be upgraded to higher speeds, or replaced by chips with different functionality.

[0029] The chips may be attached to the wafer by a mild adhesive to make it easier to handle and assemble the module. The adhesive could be as simple as water vapor, in which case no additional adhesive would be needed in an atmosphere that contains sufficient humidity. If more adhesion is required, then the surface could be coated with a suitable material. The adhesive may additionally have beneficial heat transfer properties, improving the thermal coupling between chip and wafer. A chip may be permanently attached to the wafers if there is no advantage to removing it in a particular application.

[0030] The wafers are pressed firmly together for operational use of the integrated circuit assembly module. Pressing the wafers together brings the active side of the face-to-face chips close enough to allow for capacitive communication

between the chips. The alignment structures for the chips are created with sufficient clearance to allow the chip surfaces to abut when the wafers are pressed together.

[0031] The wafers are polished to be flat. The chips may have non-uniform heights. For instance the center portion of the chips may have additional layers laminated to them, increasing the chip thickness. However, the chip surfaces are flat where the chips on the two layers contact each other. Thus, this contact surface is planar across the wafers.

[0032] The wafers may have an elastic layer, such as a polymer, that contacts the chip's back surface. If two chips have different heights, when an integrated circuit assembly module is assembled, the form of the touching chips would deform the elastic layer and help to make planar the top surface of the chips. If the elastic layer has adhesive properties and beneficial properties, then the layer could also act to attach the chips and facilitate improved heat transfer.

[0033] The silicon wafers have conductive interconnect layers in them to carry power, ground, and other signals. These layers are fabricated into the wafers before the wafers are polished smooth and the alignment structures are added. These conductive layers connect using spring MEMS structures that are patterned onto the wafers. When the wafers are pressed together, the spring heights and positions are such that the springs connect power, ground, and other signals to the chips. The active facing layer of chips contact the spring structures on the opposing wafer. The wafers may contain active circuitry to aid in regulating the voltage supply delivered to the chips. The conductive layers may contain voltage sensing traces that feedback to the power supplies the voltage delivered to the chip after $I \times R$ (current times resistance) loss. The power supplies can regulate their output voltage to correct for the $I \times R$ loss.

Integrated Circuit Assembly Module

[0034] FIG. 1 illustrates an integrated circuit assembly module 100 in accordance with an embodiment of the present invention. Integrated circuit assembly module 100 includes dies 102 (also referred to as “chips”). Dies 102 are semiconductor dies that may include any type of computational circuitry, including main memory, cache memory, and processors. Dies 102 are arranged active face to active face and overlapped to facilitate capacitive communication between dies.

[0035] Dies 102 are pressed together by wafers 104. Wafers 104 contain alignment structures that precisely position dies 102 to facilitate proximity communication. Note that wafers 104 can be silicon substrates that are comprised of the same material as dies 102. (They can alternatively be comprised of other materials that are not silicon.) One of the two wafers in wafers 104 extends than the other in one dimension to allow for perimeter chips at the edges of integrated circuit assembly module 100 to extend beyond the opposite wafer. The exposed portions of the perimeter chips allow for external I/O connections via external I/O connectors 110. These external I/O connectors 110 can be either electrical, optical, and any other type of connectors that can be used the transfer data.

[0036] On the opposite sides of wafers 104 from dies 102 are heat removal devices 106, which may include any standard heat removal devices such as heat sinks with fans. In one embodiment of the present invention, heat removal devices 106 can be made of a graphite foam (or other material, such as a metal) with fin-like structures that transfer heat to a liquid that is pumped through the graphite foam.

[0037] Coupled to the opposite side of heat removal devices 106 from wafers 104 are power supplies 108. Power supplies 108 are standard integrated

circuit power supplies, and are attached to wafers 104 via power connectors 112. Note that FIG. 1 is drawn for illustration purposes only and is not drawn to scale.

Integrated Circuit Assembly Module with Power Connections

[0038] FIG. 2 illustrates a portion of an integrated circuit assembly module 100 with power connections in accordance with an embodiment of the present invention. Integrated circuit assembly module 100 contains metal planes that provide power 207 and ground 208 to dies 102 within integrated circuit assembly module 100, such as dies 204, 205, and 206. Since dies 204 to 206 are arranged with their active circuitry side facing each other, it is beneficial to connect power 207 and ground 208 to each chip from the wafer on the opposite side. For example, power 207 and ground 208 in wafer 202 can be attached to chip 204 via MEMS spring contacts 210. Similarly, power and ground for dies 205 and 206 can be provided through MEMS spring contacts from wafer 203.

[0039] In one embodiment of the present invention, integrated circuit assembly module 100 is a complete self-contained system containing both memory and processor chips. In this embodiment, external I/O connectors 110 can connect integrated circuit assembly module 100 to other computers or to other integrated circuit assembly modules. Note that FIG. 2 is not drawn to scale.

Proximity Communication Overlap Patterns

[0040] FIG. 3 presents various arrangements of chips to facilitate proximity communication between chips. Each patterns offers a different amount of overlap between chips. In some instances it might be better to maximize the overlap between chips, such as in herringbone 302 and mosaic 306, while in other instances it might be better to maximize space between chips to facilitate heat removal, such as in checkerboard 304 and tight checkerboard 308. Various

tradeoffs exist with each pattern as well. More overlapping connections between chips might result in better bandwidth between chips, but conversely, more of the chip is being used for chip-to-chip communication and less of the chip is being used for other functions of the chip. Specific chip arrangements might prove to be optimal for specific types of chips as well. For instance, an integrated circuit module that contains only memory chips might utilize herringbone 302 as an optimal pattern, while an integrated circuit module that contains memory and processor chips might utilize checkerboard 304 as an optimal pattern.

[0041] The foregoing descriptions of embodiments of the present invention have been presented for purposes of illustration and description only. They are not intended to be exhaustive or to limit the present invention to the forms disclosed. Accordingly, many modifications and variations will be apparent to practitioners skilled in the art. Additionally, the above disclosure is not intended to limit the present invention. The scope of the present invention is defined by the appended claims.